



Audio Engineering Society

Convention Paper 7578

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Individual Subjective Preferences for the Relationship between SPL and Different Cinema Shot Sizes

Roberto Muñoz Soto¹, Manuel Recuero López², Diego Durán Blanc¹, and Manuel Gazzo¹

¹ Universidad Tecnológica de Chile INACAP, Santiago, Chile.
rmunozs@inacap.cl

² Grupo de Investigación en Instrumentación y Acústica Aplicada (I2A2), Universidad Politécnica de Madrid, España.
manuel.recuero@upm.es

ABSTRACT

The main motivation for this study was to find Individual Subjective Preferences (ISP) for the relationship between SPL and different cinema shot sizes. By means of the psychophysical method of Adjustment (MA) [1], the preferred SPL for four of the most frequently used shot sizes, i.e., wide shot, long shot, medium shot, and close-up, was subjectively quantified. Also using the Constant Stimulus Method [1], the preferred difference of SPL for different combinations of the above-mentioned shot sizes was studied. The results of this study could be used to develop sound mixing criteria for audiovisual productions.

1. INTRODUCTION

According to the Audiovisual Language, the person who receives the message is conditioned by its own characteristics, properties and essence. On the other hand, the person who sends the message must be concerned about making this message understandable. The perceptive cycle of the Audiovisual Language begins its process with the following biological characteristics: the stimulus is perceived through the visual and aural physiological mechanisms, which

determine the sensate interpretations of the diverse acoustic and luminous variations of the media presented. All this is followed by a conditioned recognition, which is stored according to biological and cultural characteristics of the subjects (memory); finally, there is a response from the person who receives the message. It is important to emphasize that it is not only a single stimulus, but a systematic group of stimuli which is organized by the subject according to its contextual situation.

Therefore, the correlation between auditory and visual stimulus is critical in order to produce an enhancement

of the subject's audiovisual experience in relation to the perceived reality of what is presented.

The evidence for the interdependence between the auditory and visual senses shows that this perceptual synergy depends on the coincidence or degree of coherence between visual and auditory information presented to the subjects.

The visual and auditory perceptions do not work as isolated processes; both modalities cooperate in the improvement of people's ability and efficiency in perceiving their surrounding environment. When the auditory information is supported by coherent visual information, or when the visual information is reinforced by a coherent auditory reference, the synergistic interaction between these two modalities reinforces stimulus comprehension [2].

As is well known in audiovisual production, there are many different relative shot sizes. To categorize them, a guide provided by the BBC with the most commonly used terms was used [3]. These shot sizes represent the distance between the spectator and the character.

In any scene, several different shot sizes are used. During the direct sound recording process, the most important task of the sound engineer is to record a good sound level with a good signal to noise ratio. In other words, the level of the recorded audio is not always necessarily related to the size of the different shots.

On the other hand, one of the most important tasks of the sound engineers during the sound mix process is to adjust the levels of different elements of the sound track. The available methods and possibilities for adjusting the levels of the sound elements are very extensive. In this way, the sound mixer is able to locate any element of the soundtrack at any distance from the spectator. From a technical standpoint, and with respect to aesthetics criteria, the question is whether or not it would be advisable to change the level of sound according to changes in the sizes of shots.

The question thus proposed is: What is the level of sound, for certain shot sizes or combinations of sizes, which enhances the power of the audiovisual experience, without interfering with the "Suspension of Disbelief" concept [4]?

It is important to mention that audiovisual productions are currently distributed in standardized formats based

mostly on ITU-Recommendation-775. The 5.1 channel system has been recommended as the standard for multichannel and stereophonic sound systems, both with and without accompanying pictures [5]. In addition, the cinema electro-acoustic system must be calibrated and equalized following a defined standard. That is why the experiments were carried out in a small dubbing stage that conformed to all the above standards.

2. METHOD

2.1. Subjects

Participants were fourteen undergraduate students of different programs at Universidad Tecnológica de Chile INACAP (10 male and 4 female, mean age of 24.4 years). All subjects reported having normal hearing and normal or corrected to normal vision.

2.2. Procedures

In this study all the audiovisual sequences were recorded in digital format. The video sequences were created at a resolution of 720 x 480 pixels. The audio of the sequences was edited using Protools HD3 Audio System and Final Cut software installed on a G4 Macintosh computer.

The visual stimuli were presented on an acoustically transparent screen of 2.4 m X 4.27 m, positioned 5.6 m from the mixing position. The image size was 4.18 m X 2.35 m, with a 16:9 aspect ratio.

Auditory stimuli were presented over a speaker arrangement which follows the ITU-R BS. 775 recommendation [5], where the L C R speakers are positioned behind an acoustically transparent screen. The audio system was calibrated at 85 dBC (slow rate) at -20dBFS pink noise signal, measured at the standard cinema listening position, and equalized following the "X" curve [6][7]. The experiment was carried out in a small cinema dubbing stage with dimensions of 8.26 m X 5.65 m X 3.45 m (see Fig. 1).

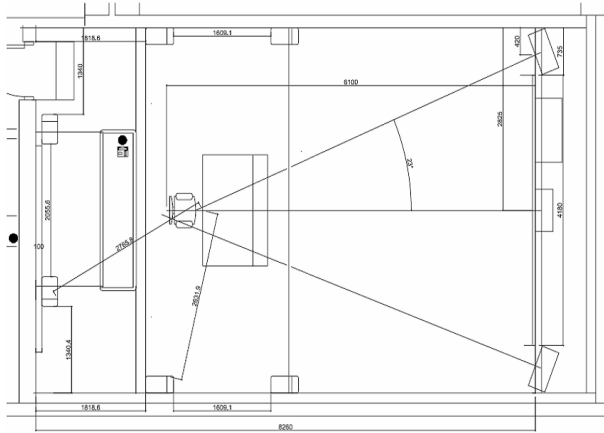


Fig. 1: Universidad Tecnológica de Chile INACAP's dubbing stage layout.

The text used was chosen taking into account the following criteria:

1. Credibility and sense of the communicative text: it is not desirable to include nonsensical words, sounds, or sentences, in order to avoid the effects of subject's comprehension and/or incomprehension on the final results of the research [8].
2. Length of the text: the sentence must be short to avoid subject fatigue.
3. Content: the text must be neutral so that it does not introduce any semantic bias to the subjects. The chosen text was "al costado del camino" ("by the side of the road"), which is a phonetically balanced sentence in Spanish.

2.3. Individual subjective preferences

2.3.1. Preferred relationship between SPL and four different cinema shot sizes

The method of adjustment was used in order to evaluate the Individual Subjective Preferences of the relationship between SPL and four different cinema shot sizes: wide shot (WS), long shot (LS), medium shot (MS), and close-up (CU) (see Fig. 2, 3, 4 and 5)



Fig. 2: Close-up



Fig. 3: Medium Shot



Fig. 4: Long Shot



Fig. 5: Wide Shot

Each of these subjects had to adjust the preferred SPL using a Control Room level potentiometer. For each shot size the subject had six trials. The sentence used, "al costado del camino", was repeated eight times per trial. The duration of each trial was 22.5 s. Each adjustment was registered on a sheet with a drawing of a graduated knob. The total time of the test was 10 minutes, including rehearsals and explanations.

2.3.2. Preferred difference of SPL for twelve different shot size combinations

In addition, twelve combinations of the four chosen shot sizes were created in order to find the preferred difference of SPL for each different shot size combination. Constant Stimulus Method with forced election was used in these cases. Ten different values of SPL difference were used in steps of 1 dB, from 1 dB to 10 dB for each couple of cinema shot sizes.

The pairs of shot sizes created are shown in the table 1.

1- CU v/s MS	2- CU v/s LS	3- CU v/s WS
4- WS v/s CU	5- LS v/s CU	6- MS v/s CU
7- MS v/s LS	8- MS v/s WS	9- LS v/s MS
10- WS v/s MS	11- LS v/s WS	12- WS v/s LS

Table 1: Pairs of shot sizes.

The sequence structures are as follows:

Each pairing of shot sizes (condition) consisted in 90 randomized trials, 9 for each one of the 10 sound volume levels differences.

The audiovisual test begins with 3 seconds of a letter from “A” to “I” followed by a beep of 1 frame of blank video and 14 black video frames. All these are followed by one and a half seconds of a number from “1” to “10” and a half of second of black video. After that, the pairs of sentences (pairs of shot sizes) were presented with duration of 5 seconds. Four seconds were given to respond. Therefore, each test of 10 sequences has a total duration 1 minute and 58 seconds. A total of 9 sets for each condition (pairs of shot sizes) were presented with a break between them of 7 seconds. The total session duration of each pair of shot sizes was approximately 20 min, including rehearsals and explanations.

On each trial, the subjects were asked to answer “yes” or “no”, as quickly as possible, if they considered coherent the SPL difference with the shot size change. They were asked to mark their answer with an “X” on a sheet with the numbers of each trial. They received instructions at the beginning of each shot size combination test, and were presented with several examples of different adjustments of volume differences (SPL).

3. MEASUREMENTS AND RESULTS

3.1. Results of the Individual Subjective Preferences of the relationship between SPL and four different cinema shot sizes

The results of each subject were averaged and the results of all of the subjects were averaged as well. The results are shown in table 2.

Shot size	Mean SPL (dB rms) C	SD (dB)
Wide Shot	69.1	7.8
Long Shot	72.4	5.3
Medium Shot	74.2	4.8
Close-Up	77.4	3.8

Table 2: Results of preferred relationship between SPL and four cinema shot sizes.

The results presented in table 2 show that:

1. The two pairs with the most different shot sizes (CU v/s WS) have the higher mean difference in dB for the mean SPL; 8.3 dB.
2. The four pairs of medium different shot sizes (WS v/s MS; LS v/s CU) have the medium mean difference for their mean SPL; 5.1 dB, and 5.0 dB respectively.
3. The rest three pairs with the least difference between them (WS v/s LS; LS v/s MS; MS v/s CU) have the least mean difference for their mean SPL; 3.3 dB, 1.8 dB, and 3.2 dB respectively.

3.2. Results of the Individual Subjective Preferences for SPL difference of different shot size combinations

For all of the 12 conditions (pairs of shot size combinations), the proportion of “coherent” responses was determined at each difference of SPL for each participant.

The method of Constant Stimuli was chosen in order to obtain the Individual Subjective Preferences. The judgments were summarized in a table (see table 3 as an example), where these values represent the percentage of the times that the SPL difference of each comparison was judged as “coherent”. The subjects were required to judge the second sentence relative to the first.

The observed distribution of responses was fitted to a Gaussian function, finding the least total error and the maximum R^2 . The percentages of “coherent” judgments for each pair of sentences were represented on a graph (figure 6).

The observed distribution was compared with a normal distribution for each participant using the Kolmogorov-Smirnov goodness-of-fit-test. All observations could reasonably have come from the specified distribution ($p > .05$) for each experiment.

The data of nine tests for each experiment were analyzed by means of a one-way ANOVA. The results of test “A”, for all experiments, were not included in the analysis. The result of the eight others test (B-I) were included because of the result of the one-way ANOVA revealed that the result of each one of them was not statistically different.

Medium Shot v/s Close-up			
SPL difference (dB)	Probability (%)	Number of trials	Subject responses
1	16	112	18
2	18	112	20
3	45	112	50
4	75	112	84
5	79	112	88
6	51	112	57
7	33	112	37
8	18	112	20
9	14	112	16
10	9,8	112	11

Table 3: Pair of shot sizes results.

To calculate the maximum and the 50% threshold of preferences, the interpolation of values of the psychometric response was required. To achieve this, a mathematical model given in Eq. (1) was used.

$$P_{(SPL)} = ae^{-(SPL - b)^2 / c^2} \quad (1)$$

Equation 2 was used to fit the curve to the observed data (see Figs. 6 and 7).

$$SPL = -(c\sqrt{\ln((P(SPL) - P_{\min}(SPL)/a))}) + b \quad (2)$$

$$SPL_{\max} = -(c\sqrt{\ln((P_{\max}(SPL) - P_{\min}(SPL)/a))}) + b \quad (3)$$

$$SPL_{50\%} = -(c\sqrt{\ln((P_{50\%}(SPL) - P_{\min}(SPL)/a))}) + b \quad (4)$$

The data on the Table 3 shows that: $P_{\min} = 9.8 \%$

As a consequence of this, and by using equation 1, the function with the best fit to the data is given by $a = 69.76$; $b = 4.68$ and $c = 2.11$, which gives a $R^2 = 0.98$ and a total error of 116.3% (see Fig. 7)

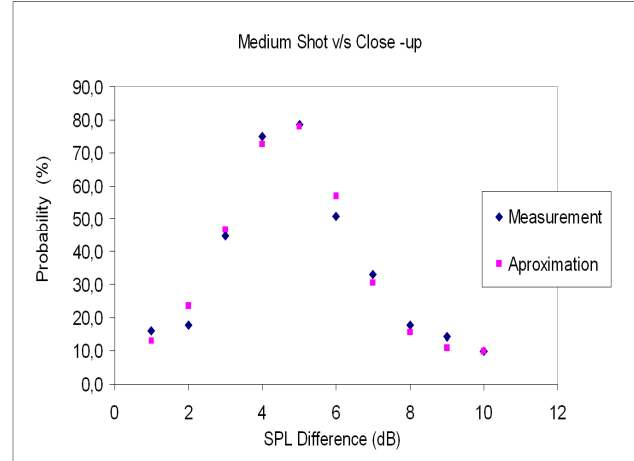


Fig 6: Values measured and their numerical approximation for “Coherent” response.

Calculating the coherence degree by using equation 4 (see Fig.7):

$$SPL_{P(50\%)} = 3.1 \text{ dB and } 6.2 \text{ dB}$$

$$SPL_{P(max)} = 4.7 \text{ dB}$$

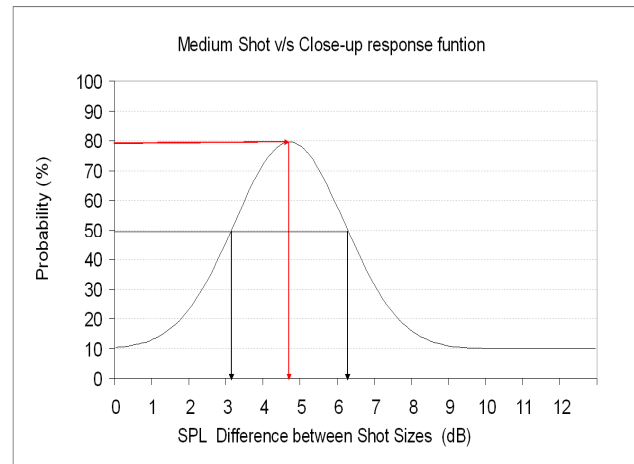


Fig. 7: Function of response.

The results of all trials are shown in table 4.

Couple	Left 50%	Mean	Right 50%
CU - WS	5.9	8.4	11.0
WS - CU	5.5	7.4	9.3
MS - WS	4.4	6.2	8.1
WS - MS	2.8	4.6	6.4
CU - LS	4.8	6.6	8.4
LS - CU	4.3	5.9	7.5
CU - MS	3.4	5.0	6.6
MS - CU	3.1	4.7	6.2
MS - LS	3.7	5.4	7.1
LS - MS	2.0	3.8	5.5
LS - WS	3.9	5.4	6.9
WS - LS	2.1	4.2	6.2

Table 4: Final results for “Coherent” response.

The two pairs with the most different shot sizes (CU v/s WS; WS v/s CU) have the higher mean difference in dB for their maximum values; 7.9 dB.

The four pairs of medium different shot sizes (CU v/s LS; LS v/s CU; WS v/s MS; MS v/s WS) have the medium mean difference for their maximum values; 5.3 dB.

The rest six pairs with the least difference between them have the least mean difference for their maximum values; 4.8 dB.

The preferred adjustments for pair of shot sizes changing from closer to farther are in average 1.2 dB over from farther to closer.

4. CONCLUSIONS

The results clearly present a preference of the audience to modify the sound pressure level when shot size changes. This was first shown when the subjects adjusted different absolute levels for different individual shot sizes. It is important to mention that the results obtained are not coincident with the divergence law (for acoustics). For instance, the difference between CU and WS mean preferred SPL was 8.3 dB (see table 2), nevertheless, the character of the WS image is 13.75 times smaller than in the CU. Hence, following the divergence law, the SPL difference would be close to 22 dB.

In the same way, a range of SPL values preferred as more coherent was obtained for each pair of shot sizes. Hence, it is recommended to modify sound pressure levels on a film sound mix when shot size changes occur. When changes occur from a small shot size to a bigger one, the recommended SPL difference necessary for coherence is greater than the SPL difference when changes occur from a big shot size to a smaller one.

This recommendation could be used by sound engineers to elaborate their own criteria for the relationship between SPL and different cinema shot sizes in order to intensify the coherence of the movie based on the preferences of the viewers.

While this paper is concerned exclusively with level, it is very important to mention that many other properties of sound give clues to distance. Equalization, reverberation, time delay between image and sound, the qualities of the other sounds in the scene, and the change of all of these with motion could be as important as level. Hence, it would be very important to study these other properties of sound in the future, under cinematographic standard mixing conditions.

5. ACKNOWLEDGMENTS

This work was developed thanks to the support of the Universidad Tecnológica de Chile INACAP and its Pérez Rosales Campus Vice-rector; Mrs. Karin Riedemann. The authors also thank Mark Berger, Adjunct Professor of Film at U.C. Berkeley and 4 time Oscar winner for best sound, for his help, useful advice and reviewing, as well to Victor Rojas and Jaime Delannoy for their efficient technical support.

6. REFERENCES

- [1] Gelfand: “Hearing, an introduction to psychological and physiological acoustics”. 3rd Ed. (Marcel Dekker Inc, New York, USA, 1990).
- [2] Woszczyk, S. Bech, V. Hansen. “Interactions between audio-visual factors in a Home Theater System: Definition of Subjective Attributes”. Paper presented at the 99th AES Convention (New York, USA, 1995).

- [3] bbc.co.uk/blast/film/tipsandtools/videoshootingguide/tool/guide_shots_1.shtml
- [4] S. Taylor. "Biographia Literaria". H.J. Jackson, Ed. (Oxford, UK 1985).
- [5] "Multichannel stereophonic sound system with and without accompanying picture". ITU-R BS. 775 (1994).
- [6] T. Holman. "5.1 Surround Sound up and Running" (Focal Press, USA, 2000).
- [7] "Cinematography – B - chain electro-acoustic response of motion-picture control rooms and indoor theatres -- Specifications and measurements". ISO 2969 (1987).
- [8] T. Soto. "Influencia de la percepción visual del rostro del hablante en la credibilidad de su voz". Tesis doctoral departamento de comunicación audiovisual y publicidad (Universidad Andrés Bello, Chile, 2000).